

# Empirical processes of iterated maps that contract on average

Olivier Durieu

Université François-Rabelais de Tours

Laboratoire de Mathématiques et Physique Théorique UMR-CNRS 7350,

Fédération Denis Poisson FR-CNRS 2964

olivier.durieu@lmpt.univ-tours.fr

March 8, 2013

## Abstract

We consider a Markov chain obtained by random iterations of Lipschitz maps  $T_i$  chosen with a probability  $p_i(x)$  depending on the current position  $x$ . We assume this system has a property of “contraction on average”, that is

$$\sum_i d(T_i x, T_i y) p_i(x) < \rho d(x, y)$$

for some  $\rho < 1$ . In the present note, we study the weak convergence of the empirical process associated to this Markov chain.

Keywords: Markov chain; Empirical processes; Weak convergence.

## 1 Introduction

The study of limit behavior of partial sums of random variables has a long time interest in dynamical systems or Markov chains theory. Existence of (attractive) invariant measure, law of large numbers, central limit theorem, almost sure invariance principle are results that characterize the statistical properties of such systems. In this note, we study the limit behavior of empirical processes, associated to some random iterative Lipschitz models. The analyze of empirical processes convergence is also of interest in dynamical systems in order to derive statistical tests, as Kolmogorov-Smirnov test. See for example [3] where expanding maps of the interval are considered. The iterative Lipschitz models that we study are described in Sections 2 and 3. Some background on empirical processes is briefly recalled in Section 4. In Sections 5 and 6, we show how results of [5] can be applied to get an empirical central limit theorem for the iterative Lipschitz models considered here.

## 2 Iterated Lipschitz maps

Let  $\mathcal{X}$  be a locally compact metric space, with a countable basis. Denote by  $d$  the metric on  $\mathcal{X}$ . Let  $T_i$ ,  $i \geq 0$ , be a sequence of Lipschitz maps from  $\mathcal{X}$  to  $\mathcal{X}$  and  $p_i$ ,  $i \geq 0$ , be a sequence of Lipschitz functions from  $\mathcal{X}$  to  $[0, 1]$  such that for all  $x \in \mathcal{X}$ ,  $\sum_{i \geq 0} p_i(x) = 1$ . We will consider the Markov chain with state space  $\mathcal{X}$  and transition probability  $P$  given by

$$P(x, A) = \sum_{i \geq 0} p_i(x) 1_A(T_i x),$$

for all  $x \in \mathcal{X}$  and all Borel subset  $A$  of  $\mathcal{X}$ . We also denote by  $P$  the Markov operator defined for measurable functions  $f$  on  $\mathcal{X}$  by

$$Pf(x) = \int_{\mathcal{X}} f(y) P(x, dy) = \sum_{i \geq 0} p_i(x) f(T_i x).$$

We assume the maps  $T_i$  contract on average, that is, there exists  $\rho \in (0, 1)$  such that for all  $x, y, z \in \mathcal{X}$ ,

$$\sum_{i \geq 0} d(T_i x, T_i y) p_i(z) < \rho d(x, y). \quad (1)$$

The case of constant  $p_i$  has been studied in [7], [2] (with applications to image encoding) or [8]. In particular, an empirical CLT has been proved in [13]. For variable  $p_i$ , such systems have been considered in [6], [9] (with applications in learning models), [1] (existence of invariant measure), [10] (clt), [11] (Berry-Essen bounds) or [12] (AISP). A lot of concrete examples, that we do not present in this short note, can be found in the articles cited above.

For our general setting, we need the following extra assumptions:

$$\sup_{x, y, z \in \mathcal{X}, y \neq z} \sum_{i \geq 0} \frac{d(T_i y, T_i z)}{d(y, z)} p_i(x) < +\infty. \quad (2)$$

$$\sup_{x, y \in \mathcal{X}} \sum_{i \geq 0} \frac{d(T_i y, x_0)}{1 + d(y, x_0)} p_i(x) < +\infty \text{ for some } x_0 \in \mathcal{X}. \quad (3)$$

$$\sup_{x \in \mathcal{X}} \sum_{i \geq 0} \frac{d(T_i x, x_0)}{1 + d(x, x_0)} \sup_{y, z \in \mathcal{X}, y \neq z} \frac{|p_i(y) - p_i(z)|}{d(y, z)} < +\infty \text{ for some } x_0 \in \mathcal{X}. \quad (4)$$

Remark that these three conditions are trivially satisfied when the family of maps  $T_i$  is finite.

We will also assume that, for any  $x, y \in \mathcal{X}$ , there exist sequences of integer  $(i_n)_{n \geq 1}$  and  $(j_n)_{n \geq 1}$  such that, for some  $x_0 \in \mathcal{X}$ ,

$$d(T_{i_n} \circ \dots \circ T_{i_1} x, T_{j_n} \circ \dots \circ T_{j_1} y) (1 + d(T_{j_n} \circ \dots \circ T_{j_1} x, x_0)) \xrightarrow{n \rightarrow +\infty} 0, \quad (5)$$

with  $p_{i_n}(T_{i_{n-1}} \circ \dots \circ T_{i_1} x) \dots p_{i_1}(x) > 0$  and  $p_{j_n}(T_{j_{n-1}} \circ \dots \circ T_{j_1} y) \dots p_{j_1}(y) > 0$  for all  $n \geq 1$ . Remark that this assumption is satisfied when (1)–(4) hold and the  $p_i$  are all strictly positive.

### 3 Spaces of Lipschitz functions with weights

We will introduce some Banach spaces, on which the operator  $P$  acts with good spectral properties. Let  $\alpha, \beta$  be non-negative real numbers. We denote by  $\mathcal{H}_{\alpha, \beta}$  the space of continuous functions from  $\mathcal{X}$  to  $\mathbb{R}$  such that  $\|f\|_{\alpha, \beta} = N_\beta(f) + m_{\alpha, \beta}(f) < +\infty$ , where

$$N_\beta(f) = \sup_{x \in \mathcal{X}} \frac{|f(x)|}{1 + d(x, x_0)^\beta} \quad \text{and} \quad m_{\alpha, \beta}(f) = \sup_{x, y \in \mathcal{X}, x \neq y} \frac{|f(x) - f(y)|}{d(x, y)^\alpha (1 + d(x, x_0)^\beta)}$$

for some fixed  $x_0 \in \mathcal{X}$ . The following statement is easily verifiable.

**Lemma 1.** *If  $f, g \in \mathcal{H}_{\alpha, \beta}$  with  $\|f\|_\infty = N_0(f) \leq 1$  and  $\|g\|_\infty \leq 1$ , then  $f, g \in \mathcal{H}_{\alpha, \beta}$  and  $\|fg\|_{\alpha, \beta} \leq \|f\|_{\alpha, \beta} + \|g\|_{\alpha, \beta}$ .*

The interest of introducing these Banach spaces is the following spectral decomposition of the Markov operator which is due to Peigné [10].

**Theorem 1** (Peigné [10]). *Assume (1)–(5) hold. Let  $\alpha, \beta \in (0, 1/2)$  with  $\alpha < \beta$ . Then*

1.  $P$  operates on  $\mathcal{H}_{\alpha, \beta}$ .
2. There exists an attractive  $P$ -invariant probability measure  $\nu$  which admits a moment of order one, i.e. for all  $x_0 \in \mathcal{X}$ ,  $\int_{\mathcal{X}} d(x, x_0) \nu(dx) < \infty$ .
3. There exists a bounded operator  $Q$  with spectral radius strictly less than 1 such that  $P = \nu + Q$ , with  $\nu Q = Q\nu = 0$ .

As a first consequence, Peigné obtained a central limit theorem for the process  $(f(X_k))_{k \geq 0}$ , where  $f$  is any bounded Lipschitz function and  $(X_k)_{k \geq 0}$  is the Markov chain starting at a fixed point  $x$  with transitions given by  $P$ . Another consequence is the almost sure invariance principle which was proved by Walkden [12]. As a result, the functional central limit theorem and the law of iterated logarithm hold. Here we address the question of a limit theorem for the empirical process associated to the Markov chain starting with distribution  $\nu$ .

### 4 Empirical processes

For a stationary process  $(X_k)_{k \geq 0}$  and a class  $\mathcal{F}$  of measurable functions from  $\mathcal{X}$  to  $\mathbb{R}$  which are uniformly bounded, we define the empirical process  $(U_n(f))_{f \in \mathcal{F}} \in \ell^\infty(\mathcal{F})$  by

$$U_n(f) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} (f(X_i) - \mathbb{E}(f(X_0))).$$

The classical cases, which were studied first, correspond to the choice  $\mathcal{X} = [0, 1]$  and  $\mathcal{F} = \{1_{(-\infty, t]} : t \in [0, 1]\}$ , or more generally,  $\mathcal{X} = \mathbb{R}^d$  and  $\mathcal{F} = \{1_{(-\infty, t]} :$

$t \in \mathbb{R}^d\}$  for some  $d \geq 1$ . For these particular cases, results of [4] can be used to derive a central limit theorem for  $(U_n(1_{(-\infty, t]}))_{t \in \mathbb{R}^d}$ . More recently, Dehling, Durieu and Tusche [5] established some results for the abstract case of  $\mathcal{F}$ -indexed empirical processes. They involve the following definition of bracketing number: If  $l, u : \mathcal{X} \rightarrow \mathbb{R}$  are two functions satisfying  $l(x) \leq u(x)$  for all  $x \in \mathcal{X}$ , the bracket  $[l, u]$  is defined by  $[l, u] := \{f : \mathcal{X} \rightarrow \mathbb{R} : l \leq f \leq u\}$ . Given  $\varepsilon, A > 0, r \geq 1$ , a subset  $\mathcal{G}$  of some Banach space with norm  $\|\cdot\|$ , and a probability law  $\mu$ , we call  $[l, u]$  an  $(\varepsilon, A, \mathcal{G}, L^r(\mu))$ -bracket if  $l, u \in \mathcal{G}$  and

$$\|u - l\|_r \leq \varepsilon, \quad \|u\| \leq A, \quad \|l\| \leq A,$$

where  $\|\cdot\|_r$  denotes the  $L^r(\mu)$ -norm. Now, for a class of measurable functions  $\mathcal{F}$ , the bracketing number  $N(\varepsilon, A, \mathcal{F}, \mathcal{G}, L^r(\mu))$  is defined as the smallest number of  $(\varepsilon, A, \mathcal{G}, L^r(\mu))$ -brackets needed to cover  $\mathcal{F}$ .

## 5 Main results

We consider the space  $\mathcal{L} = \mathcal{H}_{1,0}$  of bounded Lipschitz functions, equipped with the norm  $\|\cdot\|_{\mathcal{L}} = \|\cdot\|_{1,0}$ . Under an assumption on the bracketing number which allows to approximate functions of the class  $\mathcal{F}$  by functions of the spaces  $\mathcal{L}$ , we can derive the following empirical central limit theorem.

**Theorem 2.** *Assume (1)–(5) hold. Consider the Markov chain  $(X_k)_{k \geq 0}$  starting with the  $P$ -invariant distribution  $\nu$  and transition  $P$ . Let  $\mathcal{F}$  be a class of functions from  $\mathcal{X}$  to  $\mathbb{R}$ , uniformly bounded. If there exist  $\gamma > 1$ ,  $C > 0$ ,  $r \in (1, 2)$ , and  $c \geq 0$  such that*

$$\int_0^1 \delta^c \sup_{\varepsilon \geq \delta} N(\varepsilon, \exp(C\varepsilon^{-\frac{1}{\gamma}}), \mathcal{F}, \mathcal{L}, L^r(\nu)) d\delta < \infty$$

*then the empirical process  $(U_n(f))_{f \in \mathcal{F}}$  associated to  $(X_k)_{k \geq 0}$  converges in distribution in the space  $\ell^\infty(\mathcal{F})$  to a tight centered Gaussian process  $(W(f))_{f \in \mathcal{F}}$  with covariances*

$$\text{Cov}(W(f), W(g)) = \sum_{k=0}^{\infty} \text{Cov}(f(X_0), g(X_k)) + \sum_{k=1}^{\infty} \text{Cov}(f(X_k), g(X_0)). \quad (6)$$

This theorem can be applied for several classes of functions  $\mathcal{F}$ . Examples of such classes, as indicators of balls or ellipsoids, can be found in [5]. Let us state the result corresponding to the classical empirical process indexed by left infinite rectangles in  $\mathbb{R}^d$ .

**Theorem 3.** *Assume (1)–(5) hold and  $\mathcal{X} = \mathbb{R}^d$ . Consider the Markov chain  $(X_k)_{k \geq 0}$  starting with the  $P$ -invariant distribution  $\nu$  and transition  $P$ . If there exist  $\gamma > 1$  and  $C > 0$  such that the distribution function  $F$  of  $\nu$  satisfies*

$$w_F(\delta) \leq C|\log(\delta)|^{-\gamma},$$

then the empirical process  $(U_n(1_{(-\infty, t]}))_{t \in \mathbb{R}^d}$  converges in distribution in the Skorohod space  $D(\mathbb{R}^d)$  to a tight centered Gaussian process  $(W(t))_{t \in \mathbb{R}^d}$  with covariances as in (6).

## 6 Proofs

*Proof of Theorem 2.* We will apply Theorem 1.1 of [5]. First, as a consequence of the spectral decomposition of Peigné's theorem, we have the central limit theorem for functions of  $\mathcal{L}$  (see [10]), i.e. for all  $f \in \mathcal{L}$ ,

$$\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} (f(X_k) - \mathbb{E}(f(X_0))) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(f)), \quad (7)$$

where  $\sigma^2(f) = \text{Var}(f(X_0)) + 2 \sum_{k=1}^{\infty} \text{Cov}(f(X_0), f(X_k))$ . Now, set  $s = \frac{r}{r-1} \in (2, +\infty)$ ,  $\beta = 1/s$  and choose  $\alpha < \beta$ . From Theorem 1, we infer that there exist a constant  $C > 0$  and  $\theta \in (0, 1)$  such that for all  $f \in \mathcal{H}_{\alpha, \beta}$ ,

$$\|P^n f - \nu(f)\|_{\alpha, \beta} \leq C \|f\|_{\alpha, \beta} \theta^n. \quad (8)$$

We will deduce a multiple mixing property for  $\mathcal{L}$  functions. Fix a positive integer  $p$ . All along the proof,  $C$  always denotes a positive constant which does not depend on the choice of the functions  $f$ , but its value may change. Observe that for each  $f \in \mathcal{H}_{\alpha, \beta}$ ,

$$\begin{aligned} \|f\|_s^s &= \int_{\mathcal{X}} |f(x)|^s \nu(dx) \\ &\leq \int_{\mathcal{X}} \left( \frac{|f(x)|}{1 + d(x, x_0)^\beta} \right)^s \left( \frac{1 + d(x, x_0)^\beta}{(1 + d(x, x_0)^\beta)^\beta} \right)^s (1 + d(x, x_0)) \nu(dx) \\ &\leq C N_\beta(f)^s \end{aligned}$$

since  $\frac{1 + d(x, x_0)^\beta}{(1 + d(x, x_0)^\beta)^\beta}$  is bounded and  $\nu$  has a finite moment of order 1. Thus for all  $f \in \mathcal{H}_{\alpha, \beta}$ ,  $f \in L^s(\nu)$  and

$$\|f\|_s \leq C \|f\|_{\alpha, \beta}. \quad (9)$$

Moreover, for each  $f \in \mathcal{L}$  with  $\|f\|_\infty \leq 1$ , we have  $f \in \mathcal{H}_{\alpha, \beta}$  and  $\|f\|_{\alpha, \beta} \leq 3 \|f\|_{\mathcal{L}}$ . For all  $i_0 < i_1 < \dots < i_p$  in  $\mathbb{N}$ , by Hölder's inequality and (9), we have

$$\begin{aligned} &|\text{Cov}(f(X_{i_0}) \dots f(X_{i_q}), f(X_{i_{q+1}}) \dots f(X_{i_p}))| \\ &\leq \mathbb{E}(|f(X_{i_0}) \dots f(X_{i_q})|^r)^{\frac{1}{r}} \\ &\quad \times \mathbb{E}(|\mathbb{E}(f(X_{i_{q+1}}) \dots f(X_{i_p}) | X_{i_{p-1}}) \dots | X_{i_q}) - \mathbb{E}(f(X_{i_{q+1}}) \dots f(X_{i_p}))|^s)^{\frac{1}{s}} \\ &\leq \|f\|_r \|P^{i_{q+1}-i_q} g - \nu(g)\|_{\alpha, \beta}, \end{aligned} \quad (10)$$

where  $g = f P^{i_{q+2}-i_{q+1}}(f \dots P^{i_p-i_{p-1}} f)$ . Since for all  $n$ ,  $P^n f \in \mathcal{H}_{\alpha, \beta}$  and  $\|P^n f\|_\infty \leq \|f\|_\infty \leq 1$ , Lemma 1 shows that  $g \in \mathcal{H}_{\alpha, \beta}$ . We can apply (8) to get

$$\|P^{i_{q+1}-i_q} g - \nu(g)\|_{\alpha, \beta} \leq C \|g\|_{\alpha, \beta} \theta^{i_{q+1}-i_q}. \quad (11)$$

Further, the spectral radius of  $P$  on  $\mathcal{H}_{\alpha,\beta}$  is 1, thus there exist a constant  $C$  such that  $\|P^n h\|_{\alpha,\beta} \leq C\|h\|_{\alpha,\beta}$  for all  $n$  and all  $h \in \mathcal{H}_{\alpha,\beta}$ . Again, by Lemma 1, there exists a constant which is independent of  $f$  such that

$$\|g\|_{\alpha,\beta} \leq C\|f\|_{\alpha,\beta} \leq C\|f\|_{\mathcal{L}}. \quad (12)$$

Thus, by (10), (11), (12), we get

$$|\text{Cov}(f(X_{i_0}) \dots f(X_{i_q}), f(X_{i_{q+1}}) \dots f(X_{i_p}))| \leq C\|f\|_r \|f\|_{\mathcal{L}} \theta^{i_{q+1}-i_q}.$$

Then  $(X_k)_{k \geq 0}$  has a multiple mixing property with exponential decay on the space  $\mathcal{L}$ . Applying Theorem 4 of [4], we obtain the following moment bound: for all  $f \in \mathcal{L}$  with  $\|f\|_{\infty} \leq 1$ ,

$$\mathbb{E} \left( \left( \sum_{k=0}^{n-1} (f(X_k) - \mathbb{E}(f(X_k))) \right)^{2p} \right) \leq C \sum_{i=0}^p n^i \|f\|_r^i \log^{2p-i} (\|f\|_{\mathcal{L}} + 1). \quad (13)$$

Therefore, thanks to (7) and (13), Theorem 1.1 of [5] can be applied. The covariance structure (6) of the limit process can be derived from the expression of  $\sigma^2(f)$  and (8).  $\square$

*Proof of Theorem 3.* By assumption there exists  $r \in (1, 2)$  such that  $\gamma > r$  and (3) holds. By Proposition 4.1 of [5], if  $\mathcal{F} = \{1_{(-\infty, t]} : t \in \mathbb{R}^d\}$ , we have

$$N(\varepsilon, \exp(C\varepsilon^{-\frac{1}{\gamma}}), \mathcal{F}, \mathcal{L}, L^r(\mathcal{F})) = O(\varepsilon^{-dr})$$

and Theorem 2 completes the proof.  $\square$

## References

- [1] M. F. Barnsley, S. G. Demko, J. H. Elton, and J. S. Geronimo. Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities. *Ann. Inst. H. Poincaré Probab. Statist.*, 24(3):367–394, 1988.
- [2] Michael F. Barnsley and John H. Elton. A new class of Markov processes for image encoding. *Adv. in Appl. Probab.*, 20(1):14–32, 1988.
- [3] P. Collet, S. Martinez, and B. Schmitt. Asymptotic distribution of tests for expanding maps of the interval. *Ergodic Theory Dynam. Systems*, 24(3):707–722, 2004.
- [4] Herold Dehling and Olivier Durieu. Empirical processes of multidimensional systems with multiple mixing properties. *Stochastic Process. Appl.*, 121(5):1076–1096, 2011.
- [5] Herold Dehling, Olivier Durieu, and Marco Tusche. Empirical process of markov chains and dynamical systems indexed by classes of functions. *submitted, arXiv:1201.2256*.
- [6] Wolfgang Doeblin and Robert Fortet. Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France*, 65:132–148, 1937.
- [7] Lester E. Dubins and David A. Freedman. Invariant probabilities for certain Markov processes. *Ann. Math. Statist.*, 37:837–848, 1966.
- [8] Hubert Hennion and Loïc Hervé. *Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness*, volume 1766 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2001.

- [9] Samuel Karlin. Some random walks arising in learning models. I. *Pacific J. Math.*, 3:725–756, 1953.
- [10] Marc Peigné. Iterated function systems and spectral decomposition of the associated Markov operator. In *Fascicule de probabilités*, volume 1993 of *Publ. Inst. Rech. Math. Rennes*, page 28. Univ. Rennes I, Rennes, 1993.
- [11] M. Pollicott. Contraction in mean and transfer operators. *Dyn. Syst.*, 16(1):97–106, 2001.
- [12] C. P. Walkden. Invariance principles for iterated maps that contract on average. *Trans. Amer. Math. Soc.*, 359(3):1081–1097 (electronic), 2007.
- [13] Wei Biao Wu and Xiaofeng Shao. Limit theorems for iterated random functions. *J. Appl. Probab.*, 41(2):425–436, 2004.